基于排列熵和支持向量机的癫痫发作预测研究 *

周梦妮 [a]，崔会芳 [a]，曹　锐 [b]，王　彬 [a]，阎鹏飞 [a]，相　洁 [a†]

(太原理工大学 a. 计算机科学与技术学院; b. 软件学院, 太原 030024)

摘　要：针对癫痫发作给病人带来的巨大伤害，为临床治疗留下足够空余时间，提出一个可以预测癫痫发作的系统模型。对 21 名癫痫病人进行研究，提取具有较低算法复杂度的排列熵构成特征向量，将其输入支持向量机（support vector machine，SVM）训练出学习模型，用来识别发作期样本，利用投票机制充分考虑病人差异来判断所处状态，最终实现癫痫的实时预测。结果表明，其中 81%的发作可以提前平均 50 多分钟预测到，且具有较低的误报率。为癫痫发作预测系统的理论研究打下坚实基础。

关键词：癫痫；排列熵；支持向量机；预测

中图分类号：TP391　　doi: 10.3969/j.issn.1001-3695.2017.12.0816

Epileptic seizure prediction research based on permutation entropy and support vector machine

Zhou Mengni[a], Cui Huifang[a], Cao Rui[b], Wang Bin[a], Yan Pengfei[a], Xiang Jie[a†]

(1. *College of Computer Science & Technology*, *b. Software College*, *Taiyuan University of Technology*, *Taiyuan* 030024, *China*)

Abstract: Aiming at the great harm caused by epileptic seizures for patients and leave enough spare time for clinical treatment, the study put forword a system which can predict the seizure in advance for people with epileptic. This method based on 21 epileptic patients and extracted permutation entropy as a feature vector which has lower algorithm complexity. Then the vector was input into the support vector machine (SVM) to train a learning model and identify the ictal samples. Taking full account of patient differences, it used voting mechanism to determine the patient's state. Finally, the method realized a real-time prediction for epileptic. The results show that this method can predict 81% of the seizures with more than 50 minutes before the onset of epilepsy, and it has a low false alarm rate. The method provides a solid foundation for theoretical research of seizure prediction system.

Key words: epilepsy; permutation entropy; SVM; prediction

0　引言

癫痫是目前较为普遍的一种神经类疾病，它的发作跨各个年龄段，并且发作原因复杂，至今为止没有出现非常有效的治疗手段。除此之外，还有 30%的病人通过手术治疗和药物治疗并没有取得较好的结果，只能长期忍受随时可能出现的病痛，带来很强的焦虑情绪，给病人及家属的身心健康带来巨大的伤害[1]。因此癫痫预测已成为一个极具现实意义的研究课题。

19 世纪 70 年代开始，癫痫发作的预测研究就引起了国内外很多学者的兴趣，并且对癫痫脑电信号从动力学、非动力学角度开展了大量的研究。大量的实验表明，癫痫发作需要经历一个时间过程，并不是突然性的，它的发作存在很大的可预测性[2,3]。虽然很多研究已经成功建立癫痫发作的预测模型，但是并不能很好地应用于临床预测。

实现癫痫发作预测的关键目标之一是通过提取原始脑电信号在各个阶段的不同特征，反映大脑活动在长时间内的动态复杂变化。研究中常用的特征有时域频域特征[4,5]、Lyapunov 指数[6]、小波分析[7]、能量[8]、熵[9]等。这些特征可以分为两大类，分别是线性分析方法（如时域频域特征、能量等）和非线性分析方法（如 Lyapunov 指数、熵等）。线性分析方法通常具有较

收稿日期：2017-12-25；**修回日期：**2018-02-28　**基金项目：**国家自然科学基金资助项目（61503272，61305142，61373101，61741212）；山西省自然科学青年基金资助项目（2015021090，201601D202042）；博士后基金资助项目（2016M601287）；山西省回国留学人员科研资助项目（2016-037）；山西省重点研发计划资助项目（201603D111014）

作者简介：周梦妮（1992-），女，山西运城人，硕士研究生，主要研究方向为智能信息处理、脑信息学；崔会芳（1992-），女，河北廊坊人，硕士研究生，主要研究方向为智能信息处理、脑信息学；曹锐（1982-），男，山西吕梁人，博士，主要研究方向为智能信息处理、复杂网络、脑信息学；王彬（1983-），男，四川内江人，副教授，硕导，博士，主要研究方向为心理学、智能信息处理、脑信息学；阎鹏飞（1984-），男，山西吕梁人，讲师，博士，主要研究方向为立体视觉、虚拟现实与机器知觉；相洁（1970-），女（通信作者），山西太原人，教授，硕/博导，博士，主要研究方向为智能信息处理、脑信息学、大数据管理与分析（xiangjie@tyut.edu.cn）.

低的计算负担，但是其涉及参数少，抗噪能力差导致其在反映脑电信号动态变化时难以表现出很大的优势。相比线性分析方法，非线性分析方法在以往的研究中表现出更高的分类准确率，且具有更好的鲁棒性。科研人员希望通过提取这些特征能够在一定程度上准确反映癫痫发作从一种状态到另外一种状态的动态变化。

目前，很多研究人员为了提高癫痫预测的准确率，降低误报率，尝试使用多类特征相结合的方法或者选取一些复杂度较高的特征，这无疑会降低预测模型的时间效率，导致系统不能对即将到来的发作及时预测，难以应用于临床实践中。而癫痫病的预测系统要求具有较高的运算效率，基本可以做到实时警报，否则会延误预报时间，给病人造成巨大的伤害。脑电数据中包含大量的噪声信号，会对研究带来很大干扰，而排列熵相对其他常用的熵分析方法而言，又具有很强的抗噪能力，这一优点保证了特征提取的准确性。另外，排列熵较低的算法复杂度，带来很高的计算效率，为癫痫发作的实时预测提供一定的保证[10]。为此研究基于排列熵及支持向量机（support vector machine,SVM）的癫痫发作预报模型。

1　研究对象

使用来自德国弗莱堡医学院癫痫预测中心的公开数据集，共包含 21 例病人数据。这些病人都患难治性癫痫，其中 11 位病人在大脑神经元区有结构病变，8 位病人病灶位于海马区，其余 2 位在皮层和海马区都发生病变。数据使用 128 个信道、256 Hz 采样率（12 号病人是 512 Hz）、16 位模数转换的数码影像脑电系统采集。每位病人都采集来自颅内的 6 个电极数据，其中 3 个电极位于病灶区，剩余 3 个位于非病灶区。本文对 21 位病人进行研究，共使用 87 次发作和 168 小时的发作间期数据。

2　研究方法

本实验使用排列熵来评估大脑活动的复杂度，提取排列熵构成特征向量，将发作间期和发作期的部分数据作为训练样本，输入 SVM，通过学习得到最优分类模型，识别出发作期的样本，最终实现发作预测。实验处理流程如图 1 所示。

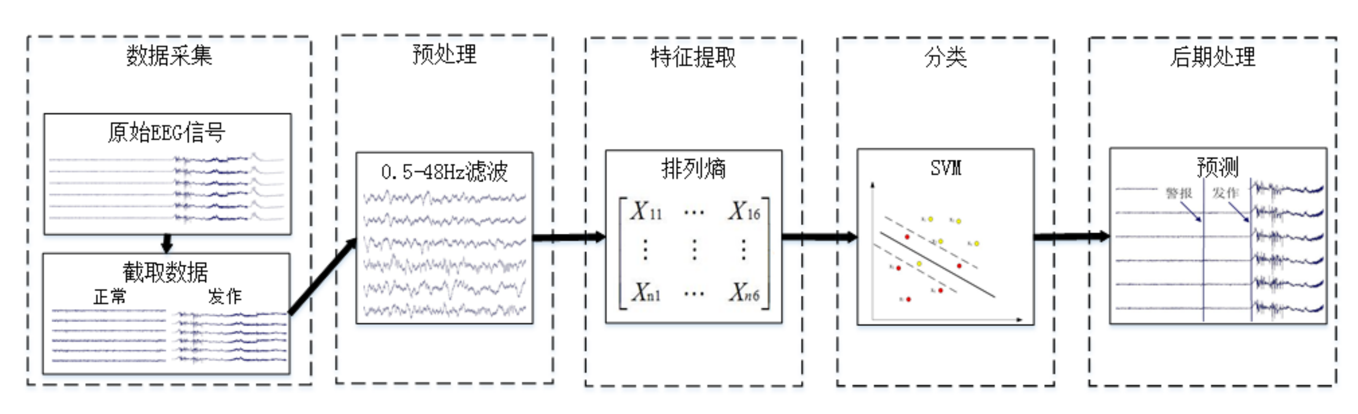


图 1　实验流程

2.1　数据准备

实验中所有病人都有 2 到 5 次包含发作阶段的数据，这些病人每次发病的持续时间较短。为保证足够的发作期样本数量，给临床治疗留出一定的干预时间，本文认为病人在发作前 30 s 已经进入发病阶段，所以发作期数据截取每次发作前 30 s 至发作结束。发作间期数据在病人完全正常期间随机选取 8 h。

2.2　预处理

由于数据采集过程中易受噪声干扰，信号中会包含脑电信号的非主要成分。本实验使用 4 阶巴特沃斯滤波器对原始信号进行 0.5~48 Hz 的带通滤波。

2.3　排列熵

排列熵是由 Bandt 等人提出的一种用来衡量一维时间序列复杂度的非线性动力学特征，现在已经广泛应用在脑电信号的研究中。因其算法复杂度低、抗噪能力强等优点[11]，相比其他的熵而言，排列熵无疑是最合适的一个选择。具体的计算方法如下：

a)已知一维时间序列{$x(i), 1\leqslant i\leqslant N$}，对该序列进行相空间重构，得到如下矩阵：

$$\begin{bmatrix} X(1) \\ \cdots \\ X(i) \\ \cdots \\ X(N-(m-1)\lambda) \end{bmatrix} = \begin{bmatrix} x(1) & \cdots & x(1+(m-1)\lambda) \\ \cdots & \cdots & \cdots \\ x(i) & \cdots & x(i+(m-1)\lambda) \\ \cdots & \cdots & \cdots \\ x(N-(m-1)\lambda) & \cdots & x(N) \end{bmatrix} \quad (1)$$

其中：m 为嵌入维数，λ 为延迟时间，$X(i)$为重构矩阵的第 i 行分量；

b)将 $X(i)$的所有元素按照升序进行排列，也就是说：

$$X(i)=\{x(i+(j_1-1)\lambda)\le x(i+(j_2-1)\lambda)\le \cdots \le x(i+(j_m-1)\lambda)\} \quad (2)$$

当出现两个元素相等时，按照 j 值从小到大来排列。

c)对每一个向量 $X(i)$都可以得到如下的符号序列：

$$S(l)=\{j_1, j_2,\cdots j_m\},(l=1,2,\cdots k, k\le m!) \quad (3)$$

其中:$j_1, j_2, \cdots j_m$ 表示重构分量每个数据所在列的位置。$S(l)$是 m 维相空间映射下 $m!$种不同的符号序列中的一种；

d)计算每种 $S(l)$出现的概率：

$$P_g(g=1,2,\cdots k) \quad (4)$$

e)按照香农熵的计算方式求得原始时间序列的排列熵，计算公式如下：

$$H_p(m) = -\sum_{g=1}^{k} P_g \ln P_g \tag{5}$$

$$H_p = H_p(m) / \ln(m!) \tag{6}$$

实验中对每个通道计算熵值时，采用 5s 的非重叠时间窗来计算。在特征提取时，嵌入维数 *m* 取 4，以便准确获得时间序列的动力学变化，保证算法的意义和有效性。一些研究表明延迟时间 *λ* 的选择对熵值的变化情况较小，这里选择 *λ=2*。

2.4 支持向量机

特征提取之后，每个 5s 的原始数据片段被表示成一个由排列熵构成的特征向量，此时需要一个具有较强分类能力的二进制分类器将此向量标记成发作期或者发作间期。大量实验表明，众多分类器中，支持向量机在二进制问题的分类中表现出很大的优势，它的优点是泛化能力强，使用过程中涉及的参数较少，比较适合对小样本数据进行分类[12,13]，且被公认为最有效的分类方法之一[14]。

支持向量机是由 Vapnik 等人在统计学习理论的基础上提出的一种机器学习方法，它的基本思想是当线性可分时，在特征空间寻找一个最优超平面，使得两类样本集离超平面的距离最大，以保证最强的泛化能力。当线性不可分时，通过引入松弛因子，使用非线性映射将低维空间的样本映射到高维空间进而转换为线性问题，加入核函数求解最优分类面。

使用支持向量机的时候，有很多核函数可以使用，比如：线性内核、多项式内核、径向基内核等。这里使用径向基核函数，因为它最适合不稳定脑电信号的分类。其中，涉及到 C 和 g 两个参数。C 是惩罚因子，g 是核参数。为了获得更好的分类模型，提高分类的准确率，将 C 和 g 的范围设置为$[2^{-8},2^{8}]$，步长为 1，然后通过网格搜索获得最优的（C,g）。

为保证实验结果的有效性，假设病人有 n 次发作，每次使用其中的 n-1 次发作和 7 小时的发作间期作为训练样本，通过 SVM 训练学习模型，对剩余的 1 次发作和 1 小时正常数据进行测试，识别发作期样本。

通常发作期的样本数远远小于发作间期的样本数，造成训练样本的严重不均衡。为了解决此问题，通过随机选择发作间期样本，将发作间期与发作期样本个数比例降为 2:1 到 8:1。SVM 可以通过为每类样本设置不同的惩罚因子来处理数据的不平衡，但是当比例严重失衡时，这一能力将大大降低甚至失效。

使用训练集形成分类模型后，对计算完排列熵之后的序列进行分类。发作期样本标记为 1，发作间期样本标记为 0。

2.5 发作预测模型

在通过 SVM 分类之后，需要对输出结果进行分析。采用 6 min 时间窗对输出结果使用投票机制，进行发作间期与发作期的样本识别。具体过程如下：

a)在特征向量输入 SVM 经过分类后，会得到相应的分类标签为

$$tg_i = \begin{cases} 1 & \text{if ictal} \\ 0 & \text{if interictal} \end{cases} (1 \le i \le x) \tag{7}$$

其中：ictal 代表发作期，用标签 1 表示。interictal 代表发作间期，用标签 0 表示。x 表示在原始时间序列以 5s 的非重叠时间窗计算排列熵之后得到的样本数量，tg_i 表示第 i 个样本的标签。

b)在此基础上，设置 6 min 的非重叠时间窗来计算这段时间内发作期标签的数量：

$$num_{ictal}[t] = \sum_{n(t-1)+1}^{nt} tg_i \tag{8}$$

这里 n 表示 72，即 6 min 的时间窗内共包含 72 个特征值，$num_{ictal}[t]$表示第 *t* 个 6 min 时间窗内的发作期样本个数。

c)使用 $num_{ictal}[t]$的值可以判断出在过去的这六分钟内信号处于发作间期还是发作期状态，具体如下：

$$state[t] = \begin{cases} 1 & if \;\; num_{ictal}[t] \ge p \\ 0 & if \;\; num_{ictal}[t] < p \end{cases} \tag{9}$$

p 值是一个阈值。当 $num_{ictal}[t]$大于等于 *p* 时，*state[t]=1*，即认为在过去的六分钟内，信号处于发作期状态；否则处于发作间期。理论上，*p* 值范围是[1,72]，实验中依据病人差异来确定。当认定该时间段内处于发作期状态时，系统发生警报。若此次警报位于发作前期，就是一次合理的发作预测，该时刻与实际癫痫发作时刻之间的时间间隔就是发作预测时间；若此次警报位于发作间期，则认为是一次误报。

2.6 模型评价指标

常见的用来衡量癫痫预测模型优劣的指标有：准确率、误报率、发作预测时间。其中，准确率指实验中预测到的次数与实际中总的发作次数之比；误报率指在非发作前期中，每小时内出现误报的次数。从模型发出警报到病人发作之间的时间间隔称为发作预测时间[15]。

3 结果与分析

3.1 排列熵在发作期与发作间期的比较

以 17 号病人为例来说明，通过提取排列熵来反映病人在发作间期、发作前期、发作期的大脑复杂度变化情况。图 2 和图 3 中均有六条折线，每条折线代表一个通道的熵值变化情况。图 2 是该病人从第二次发作前 300s 到发作完成整个阶段排列熵值的变化情况。可以看出，在前 300s（即发作前期）熵值变化基本稳定，在发作期，熵值出现显著上升后再急剧下降。当即将完成发作时，熵值又急剧上升，恢复到正常状态。其余几次发作也有类似的变化情况。图 3 是随机选取该病人在正常阶段连续 450 s 的熵值变化情况。很显然，在发作间期熵值也比较稳定。图 2 与 3 相比，发作间期和发作前期排列熵的平均水平有差异，且发作间期的均值要明显大于发作前期，说明从发作间期到发作前期再过渡到发作期，排列熵发生了复杂的动态变化。结果表明，排列熵能够很好地反映大多数癫痫病人从正常状态到发作状态再恢复到正常状态的变化过程。

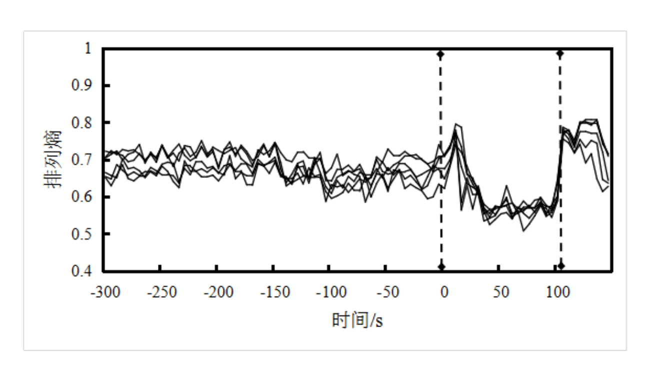


图 2　17 号病人第二次发作排列熵变化情况

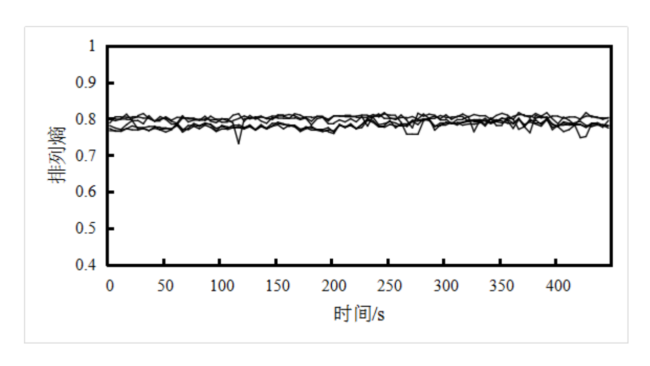


图 3　17 号病人发作间期排列熵变化情况

3.2　癫痫预测模型分析

图 4 以 19 号病人的第一次发作为例来说明预测模型的警报产生过程。原始时间序列从发作前半小时左右开始截取，直到发作完全结束。图中横轴表示时间，每相邻的两个刻度间隔表示 6min 的时间窗，灰色柱体表示 SVM 输出结果，其取值对应于主要纵坐标轴。其中，发作期用标签 1 表示，发作间期用标签 0 表示。黑色折线表示 6min 的时间窗内发作期标签的数量，取值对应于次要纵坐标轴。A 处箭头（18 min 处）表示模型发生警报所处的位置，B 处箭头（30.098 min 处）意味着癫痫开始发作。该病人的最好模型中，p 值取 12，对应于图中的黑色水平虚线。若在一个时间窗内有 12 个特征向量及以上认定为发作期状态，系统就会产生警报。

从图中可以看出，在 12-18min 时，黑色折线穿过黑色虚线，也就是说在过去的这个 6min 时间窗内，信号处于发作期状态，病人即将发病,系统在 A 处箭头（18 min 处）位置会产生警报。从产生警报到病人实际发作（B 处箭头，30.098 min 处）的时间间隔认为是发作预测时间，也就是说 18~30.098 min 之间过渡的时间称为发作预测时间，时长 12.098 min。这次警报发生在病人的发作前期，所以是一次正常警报。当警报发生在非发作前期时，则认为是一次误报。

图 5 以 19 号病人的一小时发作间期数据为例来说明没有发生误报的情况。图中各符号含义与图 4 中完全一致。18-24min 的黑色折线取值是 5，该值也是在这一小时中 6 min 时间窗内发作期标签数量的最大取值。而 p 值取 12，高于 5。说明在这一小时内没有任何一个时间窗内的发作期标签数量超过阈值，也就意味着没有任何一个 6min 的时间窗判定为发作状态。所以在这一小时内系统并没有发生警报。

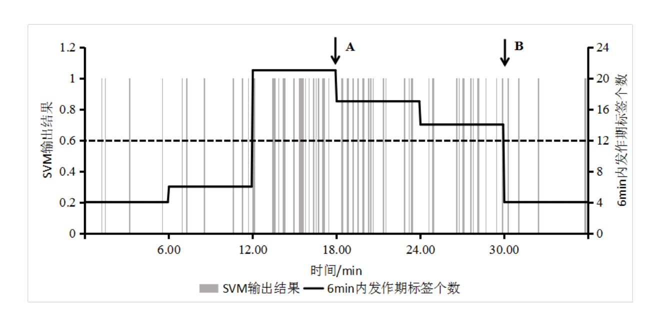


图 4　19 号病人的第一次发作警报产生过程

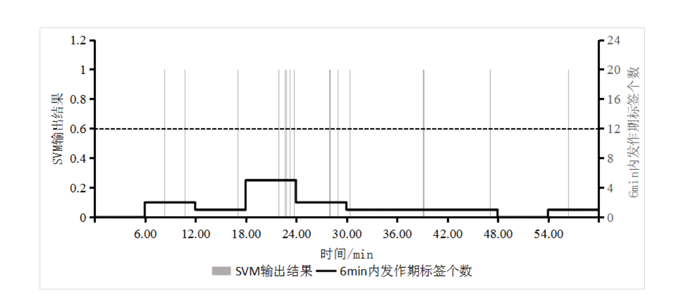


图 5　19 号病人一小时发作间期数据未产生误报

3.3　预测表现

表 1 展现了 21 个参与实验的病人评估结果。从表中可以看出，所有病人的平均预测准确率达到 81%，平均误报率为每小时 0.26，平均预测时间可达到 50.71 min。其中，部分病人取得了 100%的准确率，并且具有很低的误报率，比如 4 号、17 号等病人。这样的结果已经达到一个非常理想的状态。当然，也有个别病人取得了令人非常不满意的结果，比如 8 号病人。通过观察该病人的原始信号，发现该病人在发作期、发作前期与发作间期相比，没有出现明显差异，因此也不能提取出具有显著差异的特征，进行准确预测。还有一些病人虽然取得了相对较好的准确率，但是却有不太理想的误报率，比如 19 号病人，经过分析，本文认为主要由于发作期数据的缺乏导致。综合所有被试表现结果，误报率已经低于随机水平，也取得了不错的预测准确率。采用本方法对 21 例病人的颅内信号进行实验，表明其中 81%的发作可以平均提前 50 多分钟进行预报，且每小时的误报率仅为 0.26。

4　结束语

本实验以排列熵运算简单，时间复杂度较低，提取特征准确为基础，通过 SVM 识别出发作期的样本，使用投票机制达到癫痫发作预测的目的。与目前其他的研究相比，本方法不仅大大降低了误报率，且提前更多的时间进行预报，为临床治疗留下更充足的时间。通过分析所有病人评估结果，本研究提出的方法适用于有足够的发作次数，保证足够发作期样本，且有基本一致发作模式的癫痫病人。另外，充分考虑病人差异性使系统更好地适应临床应用。研究结果证明了通过此方法进行癫痫发作预测的可行性。

表 1　预测时间、准确率、误报率

病人编号	实际发作次数	建立模型个数	最长预测时间/min	最短预测时间/min	平均预测时间/min	准确率	误报率
1	4	32	63.89	43.89	49.88	0.69	0.56
2	3	24	58.61	13.61	39.39	0.58	0.33
3	5	40	68.61	28.61	46.19	0.55	0.08
4	5	40	62.08	33.19	49.10	1.00	0.00
5	5	40	73.96	29.58	53.47	0.60	0.28
6	3	24	58.61	41.04	49.95	1.00	0.50
7	3	24	59.03	24.86	44.54	1.00	0.17
8	2	16	73.54	3.54	36.72	0.69	0.69
9	5	40	70.63	45.49	55.60	0.98	0.13
10	5	40	78.06	43.89	61.06	1.00	0.23
11	4	32	54.79	1.53	42.58	0.66	0.22
12	4	32	51.17	24.00	34.00	1.00	0.125
13	2	16	53.82	53.82	53.82	0.50	0.19
14	4	32	76.46	9.17	53.22	0.75	0.13
15	4	32	71.46	3.82	53.50	0.38	0.38
16	5	40	66.88	6.88	47.74	0.80	0.20
17	5	40	93.19	5.42	61.81	1.00	0.10
18	5	40	78.19	14.44	56.79	0.98	0.10
19	4	32	89.17	12.10	52.25	0.97	0.81
20	5	40	87.08	2.50	55.37	0.88	0.13
21	5	40	93.13	41.60	67.94	1.00	0.00
Avg					50.71	0.81	0.26

参考文献：

[1] Bhardwaj A, Tiwari A, Krishna R, *et al.* , A novel genetic programming approach for epileptic seizure detection [J]. Computer Methods & Programs in Biomedicine, 2016, 124 (C): 2.

[2] Litt BEchauz J, Prediction of epileptic seizures [J]. Lancet Neurology, 2002, 1 (1): 22.

[3] Chaovalitwongse W A, Camey P R, Shiau D S, *et al.* , Performance of a seizure warning algorithm based on the dynamics of intracranial EEG [J]. Epilepsy Research, 2006, 72 (1): 85-7.

[4] Du X, Dua S, Acharya R U, *et al.* Classification of epilepsy using high-order spectra features and principle component analysis [J]. Journal of Medical Systems, 2012, 36 (3): 1731-1743.

[5] Gajic D, Djurovic Z, Gligorijevic J, *et al.* , Detection of epileptiform activity in EEG signals based on time-frequency and non-linear analysis [J]. Frontiers in Computational Neuroscience, 2015, 9: 38.

[6] Guler I E D. Multiclass support vector machines for eeg-signals classification [J]. IEEE Trans on Information Technology in Biomedicine, 2007, 11 (2): 117.

[7] Acharya U R, Yanti R, Zheng J W, *et al.* Automated diagnosis of epilepsy using cwt, hos and texture parameters [J]. International Journal of Neural Systems, 2013, 23 (03): 1212-3827.

[8] Omerhodzic I, Avdakovic S, Nuhanovic A, *et al.* , Energy distribution of EEG signals: EEG signal wavelet-neural network classifier [J]. Computer Science, 2013, 6.

[9] Acharya U R, Fujita H, Sudarshan V K, *et al.* , Application of entropies for automated diagnosis of epilepsy using EEG signals: A review [J]. Knowledge-Based Systems, 2015, 88 (C): 85-96.

[10] Bruzzo A A, Gesierich B, Santi M, *et al.* , Permutation entropy to detect vigilance changes and preictal states from scalp EEG in epileptic patients: a preliminary study [J]. Neurological Sciences Official Journal of the Italian Neurological Society & of the Italian Society of Clinical Neurophysiology, 2008, 29 (1): 3.

[11] Bandt C B. Permutation entropy: a natural complexity measure for time series [J]. Physical Review Letters, 2002, 88 (17): 174102.

[12] Kumar Y, Dewal M L, Anand R S. Epileptic seizure detection using DWT based fuzzy approximate entropy and support vector machine [J]. Neurocomputing, 2014, 133 (8): 271-279.

[13] Nicolaou N J. Detection of epileptic electroencephalogram based on Permutation Entropy and Support Vector Machines [J]. Expert Systems with Applications, 2012, 39 (1): 202-209.

[14] Park Y, Luo Lan, Parhi K, *et al.* , Seizure prediction with spectral power of EEG using cost-sensitive support vector machines [J]. Epilepsia, 2011, 52 (10): 1761-1770.

[15] Shiao H T, Cherkassky V, Lee J, *et al.* , SVM-based system for prediction of epileptic seizures from iEEG signal [J]. IEEE Trans on Bio-Medical Engineering, 2016, 64 (5): 1011-1022.